

Summary presentations from the ZooBank Symposium: ESA Annual Meeting, Fort Lauderdale, Florida, 18 December 2005

Andrew Polaszek, *Executive Secretary, ICZN*

As announced in BZN 63(1): 6–7, a meeting to present and discuss the ZooBank registry and other aspects of animal nomenclature and taxonomy was held as a late-breaking symposium during the 2005 Annual Meeting of the Entomological Society of America. The meeting was arranged as a series of six presentations, preceded by an introduction from Frank Krell, organiser of the symposium, and followed by a panel discussion.

PowerPoint presentations are available on the ICZN website at: www.iczn.org/Fort_Lauderdale_ZB_Symposium and summaries of the presentations are given below.

ZooBank: ICZN's open-access web-based register of all new animal names and original descriptions

Andrew Polaszek (*Executive Secretary, ICZN, London, U.K.*)

Why do we need ZooBank?

Descriptions of new animal species and associated nomenclatural acts are currently 'hidden' in thousands of specialised journals and other publications such as monographs and CDs. This process greatly reduces the 'visibility' of animal names and nomenclatural acts. Establishing ZooBank as a mandatory register of these names will largely alleviate this problem. Not only will animal taxonomic data be freely available, but also an alerting-service targeting taxa of interest to particular user groups will be provided. Completeness of the animal species register will be achieved by having registration of new names as an ICZN requirement for availability, along with retrospective registration of existing names.

With Code-compliance built into the registration process, an opportunity to introduce unprecedented stability into zoological nomenclature is being provided. The ZooBank interface will provide automatic checking for Code-compliance, and thus prevent new homonymy, stabilise spellings, fix genders and stems, and provide stability in gender agreement. As well as increased stability, the ZooBank register will provide an opportunity for increased quality control in animal nomenclature. Current issues such as the presence or absence of type specimens, accepted categories of type depositories, the use of offensive names, auctioning of names and other ethical issues, can be dealt with by amendments to the present Code. ZooBank will enable the tracking of names and hence facilitate the correction of many problems prior to publication and name availability.

As animal taxonomy moves away from its traditional journal or monograph base towards the internet, the role of a mandatory register increases in importance. Without such a register, web taxonomy would rapidly become unmanageable, and thus ZooBank will facilitate 'true' web taxonomy – i.e. taxonomy that exists only on the internet. If web taxonomy is to become a reality, however, an effective and fair peer-review system still needs to be developed.

A possible additional benefit of ZooBank would be universal availability of descriptions. Making the inclusion of original descriptions mandatory would be very difficult to achieve, partly for reasons of current copyright laws. However, ZooBank will provide a voluntary field for original descriptions (e.g. in pdf format), with no limit on numbers of illustrations. The advantages to both authors and publishers of having these descriptions freely available, with links to the original papers, will rapidly become apparent. Several prominent life sciences publishers have already agreed to make such information available to ZooBank. The nature of ZooBank as an (eventually) mandatory name register clearly separates it from other databasing initiatives such as Species 2000, ITIS, uBIO, Zoological Record / ION and ECAT.

How will ZooBank work?

Authors or third parties will be provided with the ZooBank online registration form for the submission of registration information. The usual taxonomic fields will be included, with additional fields for Code-compliance, type depositories, gender, stem, type locality details (optional) and, as discussed above, the description and figures (optional). Registration can be both pre- and post-publication, and either primary (by the author(s)) or third-party. Registration will be based on the GUID/DOI (Globally Unique Identifier/Digital Object Identifier) system. Development of GUIDs for ZooBank will be undertaken in close cooperation with GBIF (the Global Biodiversity Information Facility) and TDWG (the Taxonomic Databases Working Group). Assigning GUIDs to animal names and taxonomic acts will have several parallels with the assignment of accession numbers to gene sequences in GenBank. Also, as with GenBank, journal editors and publishers will require authors to register new taxa and nomenclatural acts with ZooBank. As discussed above, publishers will be encouraged to allow the inclusion in ZooBank of descriptive/nomenclatural sections of published work.

During the pre-publication phase there will be a holding period during which as yet unavailable names are not openly accessible. Code-compliance checks are built into the registration process, and registration will remain free to all users. ZooBank will be kick-started by making *Zoological Record's* Index of Organism Names data available via a ZooBank portal in mid-2006. Following this, an initial prototype/proof of concept will be made available as a voluntary system for a period of nine months to monitor its uptake.

Will ZooBank be accepted?

A registration system for plants was introduced into the Botanical Code at the time of the Tokyo Botanical Congress (1993) but was not ratified at the following St Louis Congress (1999), although a voluntary registration system ran for several years. Several reasons have been put forward to explain why plant and fungal taxonomists failed to adopt mandatory registration. Firstly, botanists have far fewer names to deal with, about one tenth, compared with zoologists. Secondly, there is already a very effective universal checklist of plant names in the form of the International Plant Name Index (IPNI). Finally, a section of the botanical community was unhappy about the way in which the registration clause had been introduced into the Tokyo Code. For bacteria, a mandatory registration system has been in place since 1980. Bacterial names are considered to be validly published only if published in the

International Journal of Systematic and Evolutionary Microbiology (formerly *International Journal of Systematic Bacteriology*). Mycologists have recently introduced MycoBank, a voluntary registration system.

In order to gauge acceptance by the zoological community, a dedicated discussion list for ZooBank has been established at <http://list.afriherp.org/mailman/listinfo/zoobank-list>. It has been suggested that mandatory registration of organismal names is authoritarian, and/or imperialistic, as well as requiring extra work for taxonomists. For this reason, the development and implementation of ZooBank need to be done in as user-friendly a manner as possible, and registration needs to be made straightforward. We need to be able to demonstrate that the benefits of ZooBank far outweigh any additional effort and resources required to create it. With the cooperation of *Zoological Record* staff, at least in the initial stages of development, and provision for third party registration, much of the burden is potentially lifted from authors. However, extra resources for the development of ZooBank are clearly necessary, and a business plan is therefore being developed which will be available in mid-2006. Funding will be sought from a variety of sources, including charitable trusts, foundations, and national and international donors.

To a certain extent, the establishment and acceptance of ZooBank will depend upon the adherence of zoologists to the ICZN Code rather than adoption of any other proposed nomenclatural systems such as the Phylocode or Biocode. The experience of the last few years suggests such adoption of alternative codes extremely unlikely.

The advent of web-based taxonomy seems inevitable, and thus many of the aspects of ZooBank and the ICZN Code that are affected by traditional journal or monograph-based publication will cease to be relevant. A scenario whereby the act of registration would effectively constitute publication is clearly a strong possibility in the near future. Before that can happen, a rigorous and democratic peer-review system needs to be in place to enable solely web-based taxonomy.

Finally, the year 2008 represents 250 years – a quarter of a millennium – of Linnaean zoological nomenclature. It would be extremely timely if the ZooBank register were to be complete for retrospective registration, and up to date for new animal names, by that date.

ZooBank and Zoological Record: a partnership for success

Nigel J. Robinson (*Director, Operations & Development, Thomson Zoological Ltd, York, U.K.*)

Having served the zoological research community for almost 150 years, Zoological Record (ZR) is now the oldest continuing index to the life science literature and contains the most complete and up to date record of animal taxonomy in the World. Initiated by a group of scientists associated with the British Museum in 1864, ZR was supported by the Zoological Society of London until 1980 when BIOSIS undertook production operations as joint publisher. In 2004, BIOSIS was acquired by the Thomson Corporation. Now backed by a global organization, and despite the many changes, highs and lows, over the years, ZR still continues where many others have failed, and still has its original mission to provide a service to the scientific community, with particular reference to biodiversity and taxonomy.

Today, ZR is produced by a team of 32 staff based in York, U.K. Taxonomic indexing is carried out by graduate zoologists using highly sophisticated systems and data capture procedures developed in-house specifically for ZR. These systems produce accurate output quickly, with articles generally being processed within 2–11 days of receipt; they use form-based validation and data entry with over 100 integrated checks to ensure data quality, and have allowed ZR to build publisher relations so that comprehensive coverage can be obtained in a timely fashion. ZR has never been more accurate, complete or up to date.

While ZooBank should clearly be driven and organised by ICZN, as a partner ZR can help by contributing data and back-end processing to enable the project to become reality much earlier, with greater ongoing efficiency, and with more complete data than would otherwise be possible. Many of the requirements of ZooBank are already in place in ZR data capture and processing, or are contained within the newly enhanced and freely accessible Index to Organism Names service (www.organismnames.com). These features include alerts to new names, ability for authors to submit names and publications for inclusion, original description references, links to recent articles containing the name, links to ZR and on to full text publications, links to web resources for the name, etc. So, ZR is in an ideal position to support the ZooBank names registry project.

Given the existing coverage of ZR, and the processing already in place, it is proposed that new names published in the scientific literature are captured and indexed by ZR, validated for Code-compliance, and registered in ZooBank as part of the routine ZR processing. Working with ICZN, it would be relatively easy to check Code-compliance based on ZR data capture. Much of the information required to perform these checks is already gathered as part of normal indexing operations. Any published articles submitted by authors and publishers could be routed the same way ensuring registration and inclusion of associated biology and nomenclatural acts in ZR with minimal costs and overheads for authors, publishers and ICZN. In addition, as all years of ZR will be made available electronically during 2006, the ZR data can be used to assist in retrospective registration of existing names linking to the most comprehensive set of animal names data in the world.

As a sign of commitment to the project, Thomson plans to continue working with such organizations as the Global Biodiversity Information Facility (GBIF) and Species 2000 (the Species 2000 web servers are hosted and maintained by Thomson), and will provide support to ICZN, assisting in the development of a prototype ZooBank over the coming months. The involvement of a commercial company as a partner brings the advantage of advanced, ready built technology, and stability/longevity going forward. With links to the published literature, much as full text articles, ZR is an ideal partner to accelerate, assist and participate in the ZooBank project and we look forward to a fruitful partnership.

ZooBank and GBIF

Per de Place Bjørn (*GBIF, Copenhagen, Denmark*)

The Global Biodiversity Information Facility is a megascience facility aimed at making the world's biodiversity data freely and universally available via the internet,

and sharing primary scientific biodiversity data to benefit society, science and a sustainable future. GBIF participants currently comprise 47 country members and 31 international organisations (including ICZN). There are currently 149 GBIF data providers, serving data from 517 collections. To date, these have contributed more than 73 million specimen and observation records, and more than 500,000 species records (more than a million names), most of these from the Species 2000/ITIS Catalogue of Life project. The data include, however, a large number of unregulated names from labels and field observations, and GBIF plans to develop tools to directly serve taxonomic data to GBIF from providers.

How can GBIF collaborate with ZooBank and thereby enhance taxonomy? GBIF and TDWG are setting out to form a community around GUIDs for biodiversity data. A workshop is scheduled for early February 2006 to explore infrastructure, with the possible adoption of Life Science IDentifiers (LSIDs) or Digital Object Identifiers (DOIs). LSID's take the following format: *urn:lsid:<domainName>:<namespace>:<objectId>[:<revisionId>]*; so an LSID referring to a specimen record in the GBIF network (with identifiers assigned centrally) could take the form: *urn:lsid:gbif.net:Specimen:2706712* while a record from IPNI might be in the form: *urn:lsid:ipni.org:TaxonName:82090-3:1.1*. Clearly the ZooBank project needs will be incorporated in this process. A GBIF-hosted discussion list on GUIDs is available at: <http://wiki.gbif.org/guidwiki/>.

Other multi-disciplinary fields where GBIF is planning to have a role – and where ZooBank data will be pivotal – are the development of standards for web-wide integration of taxonomical working methods (Web-enabled taxonomy) as well as standards for web-representation of broad biological information – Species Pages.

The availability of ZooBank data will also be heightened by dissemination through the GBIF portal as GBIF will form a link between taxonomically related databases and databases about animal distribution and ecology, including data related to conservation and genomics. Naturally, new entries can be immediately available through the GBIF portal and network.

For instance, the linkage between specimens and observations and the unique ZooBank entries will largely reduce ambiguity in biodiversity science.

The integration of existing, well-curated and reviewed Global Species Databases as available through the Catalogue of Life Partnership may form a future structure to enable retrospective capture and registration of animal names.

Implementing the digital taxonomic revolution: alternative strategies for a web-based registry of taxonomic names

Richard Pyle (*Bishop Museum, Honolulu, Hawaii, U.S.A.*)

Registration, publication, and availability can be defined for our present purposes as follows: Registration is the process of entering a complete record in the ZooBank registry. Publication refers to Code-compliant published works, as defined in Chapter 3 (Articles 7–9) of the 4th Edition of the ICZN Code. An available name is a scientific name applied to an animal taxon that conforms to the provisions of the Code. Below I will present the following three scenarios relating to registration and publication, and how they affect, and are affected by, the current Code: 1. (Publication +

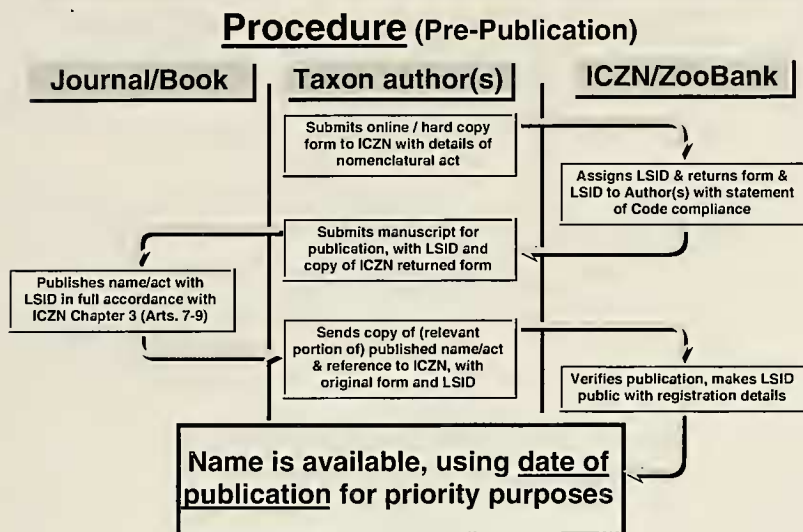
(Publication + Registration) = Availability

Figure 1

Registration) = Availability (Polaszek *et al.*, 2005; *ZooBank Technical Article*, pp. 3–5); 2. Registration = Availability (Polaszek *et al.*, 2005; *ZooBank Technical Article*, pp. 5–9); 3. Registration = Publication = Availability (Doug Yanega post to *ZooBank List*, 22 Sep 2005).

1. Publication + Registration = Availability. To be available, names and acts must be published in accordance with existing Code rules, *and* be registered. Registration can take place either before publication, or after publication. If before or within two years after publication, the date of availability is the publication date (figs. 1 & 2). If more than two years after publication, the date of availability is the registration date (fig. 3). The advantages of this scenario include relatively small changes to existing taxonomic practice, rapid implementation via an amendment to the 4th edition of the Code, the maintenance of implicit quality control via traditional publication venues, and consequently, perhaps, broader acceptance by the taxonomic community. Possible disadvantages include a somewhat complex procedure involving asynchronous publication and registration events, arbitrary time periods affecting dates of availability, and petitions to the Commission in certain special circumstances. However, given the existing complexities of the ICZN Code these procedures can hardly be considered as particularly complex. Another possible perceived disadvantage would be an ambiguous 'grey zone' between publication and registration when names and acts are 'assumed' to be available, even though technically not available until registered. Again, the probability is that most authors will register new names prior to publication, eliminating this problem entirely. While this scenario still suffers from all the complexities and ambiguities associated with traditional paper-publication entangled with nomenclatural availability, it would hardly differ from current practice, so would not really add up to an increase in complexity. Finally,

(Publication + Registration) = Availability

Procedure (Post-Publication, within 2 years)

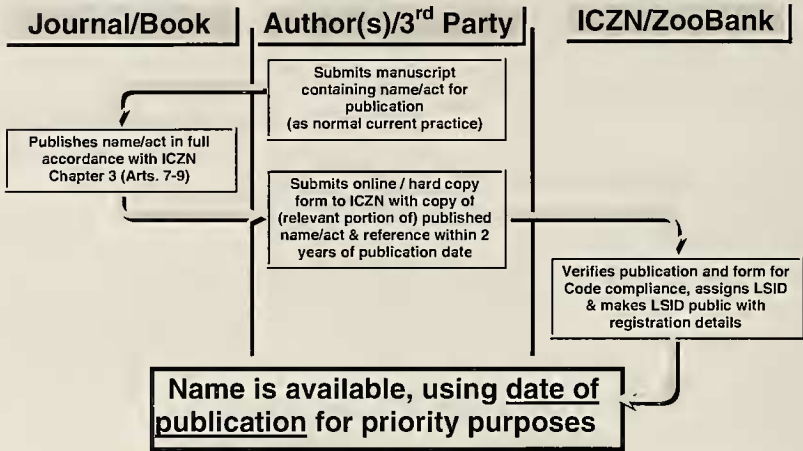


Figure 2

(Publication + Registration) = Availability

Procedure (More than 2 years Post-Publication)

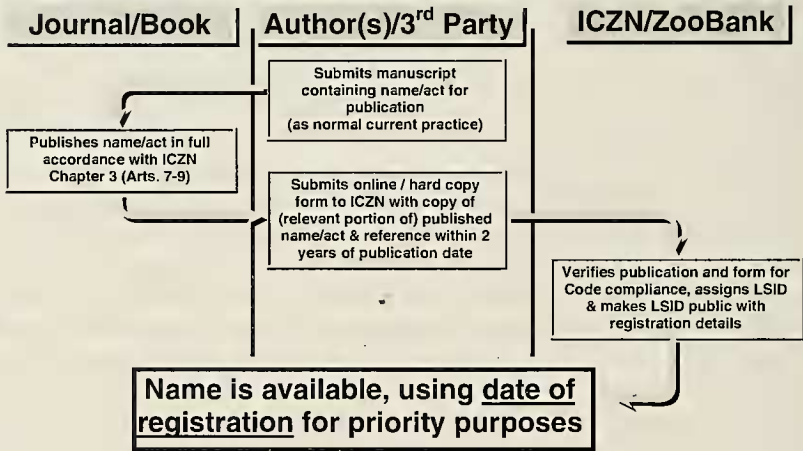


Figure 3

scenario 1 would require a (possibly extensive) increase in the active role of ICZN Secretariat staff (and associated costs) to process registration requests and verify Code-compliance before issuing GUIDs and exposing registration details to the public.

Registration = Availability

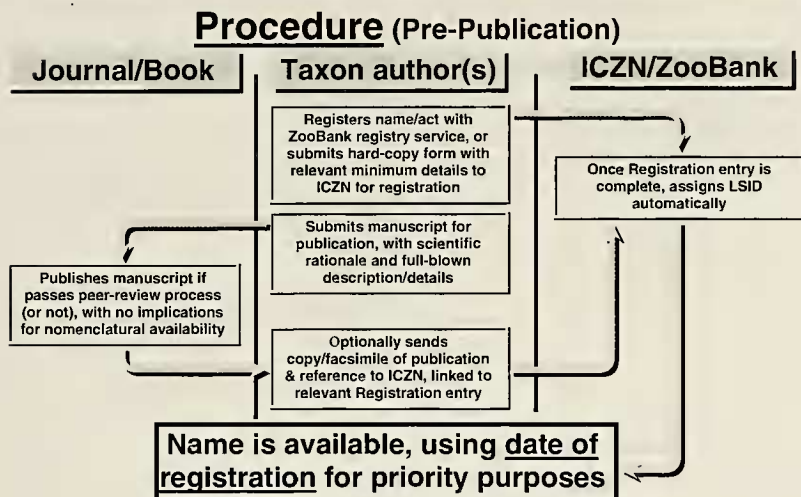


Figure 4

2. Registration = Availability. With this procedure, the process of registration itself is all that is required for availability of new names and acts. Prior or subsequent publication through traditional venues is encouraged, but is not integral to nomenclatural availability (fig. 4). Some advantages of this system would be that the legalities of nomenclatural availability and the science of taxonomy are disentangled from each other; there would be no ambiguity about dates of availability; existing complexities of nomenclatural availability of published works are moot, and only minor increases to the active role of ICZN staff (and associated costs) are envisaged. Possible disadvantages include a fundamental change to the way taxonomic names and acts are established (eliminating the publication process from the act of nomenclatural availability). However, this would not necessarily be a problem from the perspective of the taxonomists (i.e. virtually the same as scenario 1), and in fact would only require a change to the technical legality of nomenclatural availability, not necessarily any change to taxonomic practice. To implement this system, more extensive changes are also needed in the Code, such that these could probably only be implemented in a 5th Edition of the Code (perhaps 5–10 years away). However, it will probably anyway take several years to work out the details and demonstrate the feasibility via a working voluntary registration system. Another possible objection is that taxonomists would lose their primary benchmark for establishing professional status, i.e. their CVs would have fewer publications listed. Taxonomists' professional status is established by publishing articles on scientific taxonomy and classification, which would continue exactly as before; only the legalities of nomenclature would be dissociated from publications – not the science of taxonomy. While it is possible that some journals might not want to publish taxonomic descriptions if articles no longer carry the 'prestige' of establishing new names and acts in accordance with ICZN rules, it is also true that prestige in scientific publications

Registration = Publication = Availability

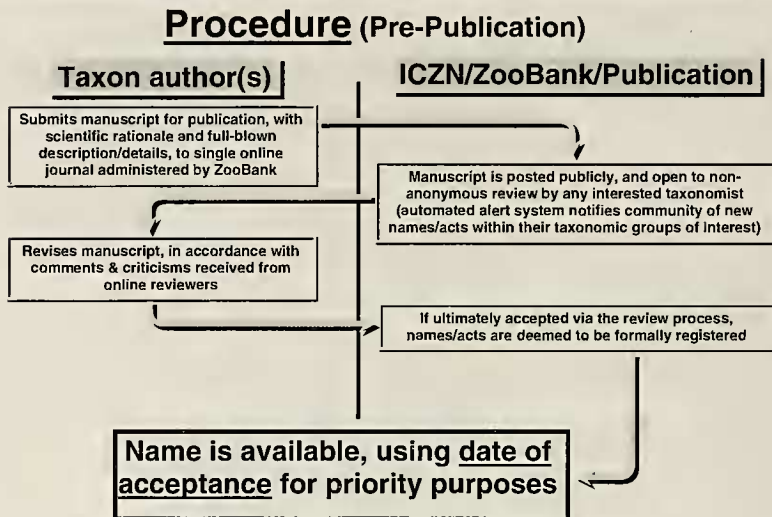


Figure 5

comes from the quality of the science content of the published articles, and not from fulfilling a legalistic technicality for nomenclatural availability. Elimination of quality control/peer review from the process of establishing new names and nomenclatural acts could also be perceived as a disadvantage, but since the Code requires neither peer review nor quality control, the scenario would be no different from the current situation. It could also be argued that the ICZN requirement for publication *de-facto* forces most names and acts through peer review anyway. The possibility that bad taxonomists (and non-taxonomists) might abuse the system by registering hundreds of bogus and unneeded names, perhaps for unscrupulous reasons (e.g. selling names for money) is also unaffected by the choice of possible scenarios – i.e. it always remains possible. The same goes for those taxonomists who might never get around to publishing the full description after the name is registered, potentially creating many names without robust taxonomic definitions.

3. Registration = Publication = Availability. Under scenario 3, the registration website, ZooBank, would host a comprehensive, edited and peer-reviewed online journal (such as *Zootaxa*) in which *all* names and acts must be published. The science of taxonomy becomes part of the nomenclatural process (by changes to the Code), and submitted manuscripts are open to non-anonymous review by any interested or concerned taxonomist (fig. 5). Major advantages of this procedure include zootaxonomic publications appearing in a single venue (as is now done for Bacteria), instead of scattered across thousands of journals, and the prevention of unscrupulous authors 'stealing' by trying to submit plagiarised work to a journal that has a faster turnaround time. All manuscripts would be examined by a large contingent of reviewers, instead of just a handful, greatly improving the reviews as well as democratizing the process. The reviews are also public, instead of anonymous, so personal grudges or biases of the reviewers are exposed to scrutiny by the whole

community. Furthermore, a dedicated nomenclatural journal would mean that the review criteria will explicitly address all necessary aspects of Code-compliance and proper nomenclature. Other advantages of an online review process include speed and openness to feedback. Above all, copyright issues would cease to be a problem. This scenario would, of course, represent a major and fundamental change to the way taxonomy is done, both in terms of legalities of nomenclature as well as for the science of taxonomy. With such major changes come particular difficulties, but the trade-off may well be worthwhile. With respect to online peer-review, it must be borne in mind that many taxonomic groups do not have many (or even any) experts who might serve as reviewers, and thus submitted manuscripts may never receive peer-review. This problem is equally true for traditional publication venues as well, but with only one 'official' taxonomic journal with potentially thousands of regular contributors and readers, there is a much better chance of finding someone who is qualified to review the manuscript. As with scenario 2, above, more extensive changes to the Code would be required, such that it could probably only be implemented in the 5th edition, perhaps 5–10 years in the future. In any case, it will probably take several years to work out the details and demonstrate the feasibility via a working voluntary registration system. It could be argued that such a system would impose a huge burden on the taxonomic community to provide peer reviews to 20,000+ new names each year, but in fact the burden would be no more than already exists. For every manuscript submitted and reviewed through the official ZooBank online journal, one fewer manuscript would be submitted to a traditional journal, so there would be no net increase in the total number of manuscripts to review. A common argument against such a scenario is that existing journals that depend on taxonomic descriptions and nomenclatural acts to fill their pages and maintain a subscriber base may be driven out of business. Since when is it the job of scientists to keep journal publishers in business? Journals exist to serve scientists, not the other way around. Criteria for determining when a submitted manuscript should be deemed 'accepted', and when (and by whom) will always be a subjective and contentious issue. This problem could be largely solved by having each manuscript assigned to an impartial 'referee' whose speciality is *outside* the particular taxon involved, and who is fully familiar with the code – serving the same role as a journal editor. Finally, the legalities of nomenclatural availability, and the subjective science of taxonomy, would, for the first time, be formally coupled under Code rules. Controversial as this sounds, it may be that a significant proportion of zoologists feel that quality control and peer review *should* be part of the Code's requirements for nomenclatural availability.

Copyright: the new taxonomic impediment

Donat Agosti* & Norman F. Johnson** (**American Museum of Natural History, New York, U.S.A. and Naturmuseum der Burgergemeinde, Bern, Switzerland; **Ohio State University, U.S.A.*)

Copyright is a set of exclusive rights granted by government for a limited time to regulate the use of a particular form, way or manner in which an idea or information is expressed. In law, an exclusive right is the power or right to perform an action in

relation to an object or other thing which others cannot perform. The law may require that a person seeks such rights through application, or it may automatically grant such rights. Exclusive rights may be granted in intellectual property law. Most governments recognize a bundle of exclusive rights in relation to creative and scientific works and property under the umbrella term 'intellectual property'. An example is copyright (Source: Wikipedia).

An example of copyright law impeding progress in taxonomy is revealed if we examine the number of publications dealing with ant species in 2003, when fewer than 5% were open-access and the remainder copyrighted.

In 2001 UNESCO produced the Universal Declaration on Cultural Diversity which reads: 'While ensuring the free flow of ideas by word and image, care should be exercised that all cultures can express themselves and make themselves known. Freedom of expression, media pluralism, multilingualism, equal access to art and to scientific and technological knowledge, including in digital form, and the possibility for all cultures to have access to the means of expression and dissemination are the guarantees of cultural diversity'.

Recent years have seen an unprecedented rise in the number of scientific and other documents being scanned, and in many cases made freely available via the internet. However, we feel that the copyright issue is still far from being addressed. In our opinion, taxonomic publications are 'legal documents', they must conform to the Codes to make the nomenclatural decisions presented valid. Thus everybody should have access to these legally binding documents. Taxonomic descriptions are also factual knowledge, that is knowledge based on direct observations. Thus, taxonomic publications, at least their descriptive parts, cannot be copyrighted, and should be open access. Species (or taxa in case of higher level revisions) descriptions can be considered the building blocks or basic data elements of taxonomic publications. They are very rich in detail. All the other elements of a publications are inferred from the analysis and synthesis of taxon descriptions. The descriptions are also the 'legal' element of the publication in compliance with the ICZN Code. Species descriptions can be further resolved into the basic units, characters in the description *sensu stricto*, and the specimen records, which are a species at a given time at a given locality, i.e. a collecting event. They could be enhanced by shared ontologies and gazetteers.

The deconstruction of text documents to, for example, XML format would be a means of placing descriptions in the public domain while circumnavigating copyright issues. Several initiatives for data domains and standards are currently being developed, including the Taxonomic Concept Schema, ABCD, DarwinCore, TaxonX and TaXMLit. Such document deconstruction projects could finally put species descriptions where they belong: firmly in the public domain, and allow third parties to build applications to mine, extract and integrate the very data-rich content of most descriptions.

Name Registration: One fewer impediment to taxonomy

James B. Woolley (*Texas A&M University, U.S.A.*)

We are currently witnessing a renaissance in systematics. Traditional approaches to collecting, specimen preparation and study, and the production of, and access to,

published work, have been revolutionized through new technologies available to taxonomists. Digital technologies have changed all the rules, and taxonomic collections, literature, expertise, digital libraries and virtual monographs need to become distributed, virtual research tools and education resources. The new taxonomy needs to be *web-based*, providing a single, global point of access; *distributed*—for example, there are currently more than 350 web sites just for Lepidoptera; *authoritative*—we need an electronic catalogue of life; *accessible* to multiple audiences and *relevant* to societal concerns such as natural resource management, invasive species, agriculture and medicine. Taxonomic publications should not be end points, but ‘version control’ devices.

However, there are currently several impediments to the implementation of this new taxonomy, including lack of funding (most funding for systematics is devoted to constructing molecular phylogenies, not taxonomy), a dearth of taxonomists, difficulties inherent in the science itself, and scattered resources in terms of both specimens and literature. Funding issues have been recognized by the USA’s National Science Foundation, with their implementation of Planetary Biodiversity Inventories, revisionary syntheses in systematics and the PEET (Partnerships for Enhanced Expertise in Taxonomy) programme. The establishment of ZooBank will go a long way towards providing centralized sources of zoological taxonomic information.

The overarching finding of the recently published ‘Atkins Report’ (Atkins *et al.*, 2003), by authors from both academia and industry, is that a new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information, and communication technology, and pulled by the expanding complexity, scope and scale of today’s challenges. The capacity of this technology has crossed thresholds that now make possible a comprehensive ‘cyberinfrastructure’ on which to build new types of scientific and engineering knowledge environments and organizations, and to pursue research in new ways and with increased efficiency. This cyberinfrastructure will be used to build more ubiquitous, comprehensive digital environments that are interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments. Such a cyberinfrastructure will include grids of computational centres, some with computing power second to none; comprehensive libraries of digital objects including programmes and literature; multidisciplinary, well-curated, federated collections of scientific data; thousands of on-line instruments and sensor arrays; convenient software toolkits for resource, discovery, modelling and interactive visualization and the ability to collaborate with physically distributed teams of people using all of these capabilities. Again, according to the Atkins Report, many contemporary projects require effective federations, distributed resources (data and facilities) and distributed, multidisciplinary expertise. Examples of ‘Virtual Science Communities’ include: National Ecological Observatory Network (NEON), National Virtual Observatory (NVO), Space Physics and Aeronomy Research Collaboratory (SPARC), Grid Physics Network (GriPhyN), Biomedical Informatics Research Network (BIRN) and National Science Digital Library (NSDL). During the last three years NSF has sponsored workshops on taxonomy, systematics, imaging and databasing. These workshops have called for a national framework for taxonomic research and natural history collections. One vision of this framework is the LINNÉ project (Legacy

Infrastructure Network for Natural Environments). LINNÉ would be a distributed, virtual taxonomic cyberlaboratory, of which each collection or taxonomic research facility is potentially a node.

Implementation of LINNÉ will modernize the national infrastructure for taxonomic research with high resolution two- and three-dimensional surface and internal scanning using computer tomography, remote-controlled digital microscopy, comprehensive digital libraries, modern collection facilities, the provision of comprehensive access to taxonomic and collections information worldwide, and new tools for education and outreach. This 'virtual research platform' will address the so-called 'big questions': What are earth's species, and how do they vary? How are species distributed in geographical and ecological space? What is the history of life on Earth, and how are species interrelated? How has biological diversity changed through space and time? What is the history of character transformations? What factors lead to speciation, dispersal and extinction?

Virtually all of the necessary technology is already in place or will be in the next few years in order to implement this vision. Many national and international activities are already underway, and initiatives such as the National Biological Information Infrastructure (NBII) are linking databases, informatics products and analytical tools for data sharing among governmental agencies, NGOs, academic institutions and industry. Similarly, GBIF, operating at the intersection of science, policy and applications, currently comprises 47 member countries, and is especially concerned with access, diversity of data, setting taxonomic standards (including the development of ECAT, the Electronic Catalogue of Life), data quality, data cleaning tools, interoperability (including GUIDs – global identifiers for specimens and collections) and collaboration. GBIF can provide critical components of cyberframework for LINNÉ, and in exchange, LINNÉ will provide data to GBIF. The SYNTHESYS initiative includes 20 European natural history museums and botanic gardens and is funded through the FPVI Integrated Infrastructure Initiative Grant. Starting in 2004, the objective of this 5-year project is to create an integrated European infrastructure for researchers in the natural sciences. 20 institutions and 11 national taxonomic facilities are involved in a two part plan comprising access and networking activities. Other initiatives include the European Network for Biodiversity Information (ENBI), the European contribution to GBIF, the CHRONOS project for the earth sciences community, the National Ecological Observatory Network (NEON) and the Natural Sciences Collections Alliance. Thus the foundations are already demonstrably in place, and the challenge is not to invent all of the necessary components *de novo*, but rather to identify what is already there, to implement the new cyberinfrastructure and integrate these components into an operational system. To do this will require that we establish a common vision and research agenda, and that we work as a community, worldwide to achieve it. This will require a change in our scientific culture necessitating an integrated, 'big-science' approach, and we need to identify common goals and work together. Other communities have done this, but there were some tough transitions. For example, particle physicists had terrible problems with career recognition and rewards with the switch to a big science paradigm. Other challenges include the fact that it will cost billions of dollars; will require Congressional and State action; it will require a unified user community, will take many years and will not be easy. However, if successfully

achieved, LINNÉ will preserve our heritage, revitalize taxonomy, and will be the most important new tool available to biologists in the 21st century.

Reference

- Atkins, D.E., Droegemeier, K.K., Feldman, S.I., Garcia-Molina, H., Klein, M.L., Messerschmitt, M.L., Messina, P., Ostriker, J.P. and Wright, M.H. (2003). Revolutionizing science and engineering through cyberinfrastructure. *Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, 1–52. NSF, Washington, D.C.